Aspect-based Sentiment Analysis

Abdul Momen Usmani

Abstract

This paper investigates various neural network architectures to enhance the performance of aspect-based sentiment analysis (ABSA). We propose three model variants using Bi-LSTM model based on aspect location: the first uses word embeddings of sentence list concatenated with aspect embeddings; the second concatenates the hidden states of the Bi-LSTM output with the aspect embeddings; the third variant with an attention mechanism, which processes the concatenated words embeddings of sentences and aspect embeddings through the Bi-LSTM model with attention. Our experiments are conducted on a customized dataset of restaurant reviews, annotated with eight aspect categories (food, service, staff, price, ambience, menu, place, and miscellaneous) and three sentiment polarities (positive, negative, neutral). Through comprehensive experimentation and analysis, we aim to identify the optimal model for sentiment classification. Our results demonstrate that the Bi-LSTM with attention mechanism achieves the highest accuracy (approx 81% for training and 66% for test data), outperforming the other two variants in training evaluations. We also did ablation study such as using Bi-RNN, BI-GRU, aspect location variants neural networks.

1 Introduction

Aspect-based sentiment analysis (ABSA) is a crucial task in natural language processing (NLP) that aims to identify sentiment towards specific aspects within a text. One important topic in Natural Language Processing (NLP) is Aspect-based Sentiment Analysis (ABSA), which aims to identify feelings expressed towards particular features in a document. By identifying sentiments associated with specific elements, such "food" or "service" in a restaurant review, ABSA delivers deeper understanding than traditional sentiment analysis, which only produces an overall sentiment score for the text. Recently, there has been a lot of promise demonstrated by neural networks, especially those that use Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, for capturing contextual information and dependencies in text. According to our findings referring to [**1**] and [**5**] we find that using models like Bi-LSTM is a very good approach as its better than methods like RNN when it comes to Vanishing Gradient and gradient explosion problem [**2**]. LSTM models consists of and Input gate (i), a Forgetting gate (f) and an Output gate (o) and also a Memory unit (C_t). These three gates described are used to record and update the Memory unit information [**1**]. The formulas for f, i, o, and C_t are as follows:

$$
\begin{aligned}
f &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) && (1)\\
i &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) && (2)\\
o &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) && (3)\\
C_t &= f_t \times C_{t-1} + i_t \times \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) && (4)\\
h_t &= o_t \times \tanh(C_t) && (5)
\end{aligned}
$$

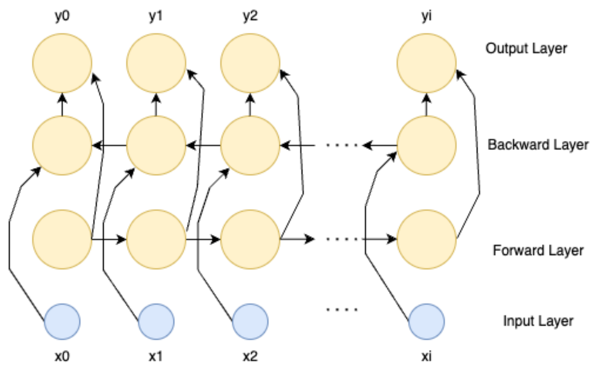

Figure 1: Bi-LSTM architecture

Each word of the sentences are represented by a pre-trained 50 dimensional vector. Padding of zero vectors of dimension same as for words are used in sentences to match the sequence length.

We followed three approaches as model variations which are as follows:

1. **Variant 1**: This approach appends aspect embeddings to word embeddings, treating the aspect as part of the sentence. The preprocessed results are then fed into the Bi-LSTM model.

2. **Variant 2**: This variation concatenates hidden layer outputs with aspect embeddings, treating the aspect as separate from the sentence to compare with the first model's approach.

3. **Variant 3**: This variation appends word and aspect embeddings, passing them through a Bi-LSTM with attention. The attention mechanism helps the model focus on the most important sections of the sentence related to the aspect.

2 Methods

Our approach involves designing three model variants to integrate aspect information differently which is shown as follows.

2.1 Variant 1

This model variant takes pre-processed sentence embeddings as input. Aspect embeddings are appended to the word embeddings in sentences, along with padding to match the sentence length. The model features a bidirectional LSTM capturing information from both directions. The final hidden states (Equation 5) from the LSTMs are concatenated and passed through the output layer for sentiment classification using a softmax function, as shown in **Figure 2** and the equation for this variant is given right below.

$$\text{hidden states } h_t = \left[h_t^f; h_t^b\right] \tag{6}$$

$$\text{sentiment} = \text{softmax}(f(h_t; w_o)) \tag{7}$$

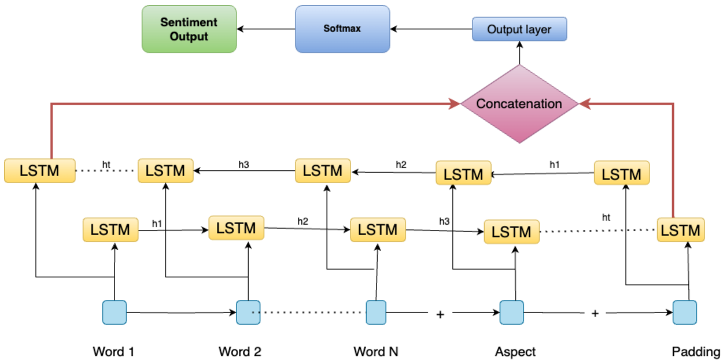

Figure 2: The architecture of the Bi-LSTM model with Aspect-augmented Input.

2.2 Variant 2

In this model variant, sentence embeddings are processed without appending aspect embeddings. The embeddings are padded with zero vectors to match the sequence length and passed through a bidirectional LSTM. The hidden states from the forward and backward LSTMs are concatenated, then the aspect vector is concatenated with it. This final vector is passed through a linear output layer, where a softmax function categorizes the sentiments (Equation 8), as shown in **Figure 3** and the equation for this variant is given right below.

$$\text{sentiment} = \text{softmax}(f((h_t; \text{aspect}); W_o)) \tag{8}$$

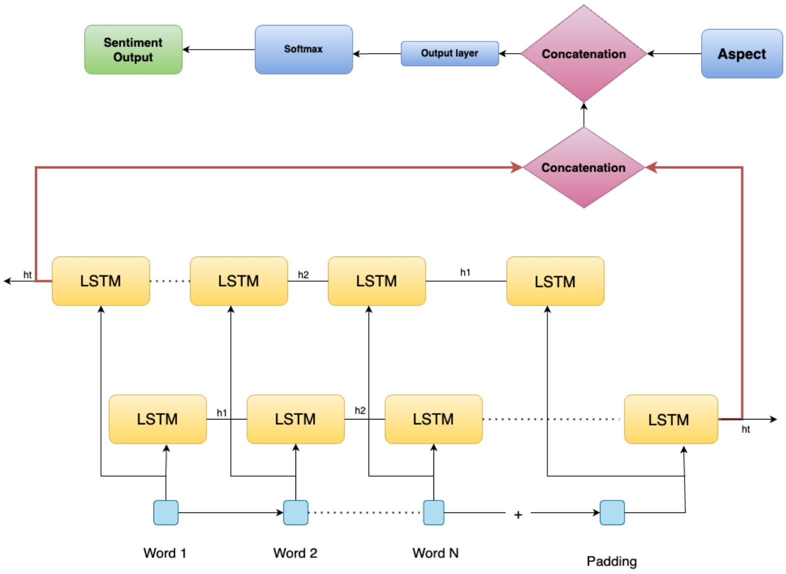

Figure 3: The architecture of the Bi-LSTM with Aspect-weighted Hidden Layers.

2.3 Variant 3

In this variant, we use an attention mechanism with a Bi-LSTM model. Sentence embeddings are preprocessed, appending aspect embeddings and padding to match sequence length. LSTM hidden states from all timestamps calculate attention weights (Equation 9) using two neural network layers followed by a softmax function. These attention scores (Equation 10) are element-wise multiplied with the LSTM hidden state outputs (Equation 11) to get the attention-weighted outputs, which are then passed to the output layers for sentiment classification, as shown in **Figure 4** and the equation for this variant is given right below.

$$\begin{aligned} &g_i = f_1(f_2(h_1; h_2; \ldots; h_t)) \\ &\quad \text{where } f_1 \text{ and } f_2 \text{ are the functions applied by respective attention layers} \end{aligned} \tag{9}$$

$$\alpha_i = \text{softmax}(g_i) \tag{10}$$

$$O_i = \alpha_i \odot h_i \tag{11}$$

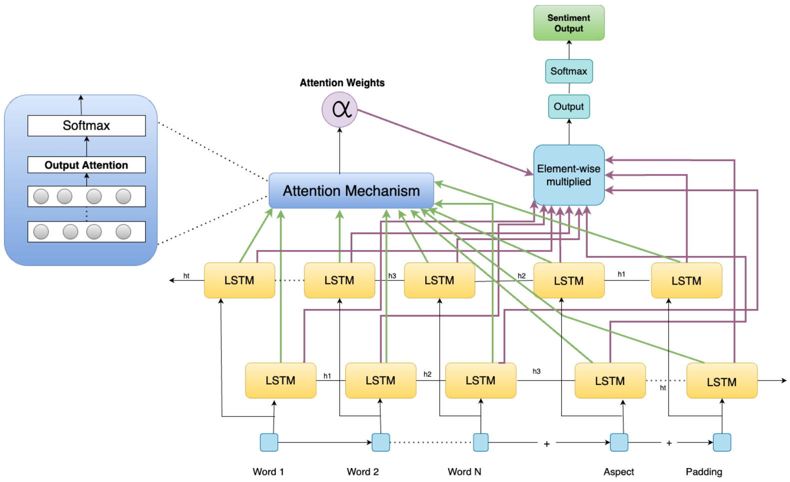

Figure 4: The architecture of the Bi-LSTM model with attention.

3 Experiments

3.1 Dataset Description

The dataset we have aims to analyze textual feedback across many forms of user experience. Together with assessment of sentiments expressed about discrete aspects of services or products, in our case its from restaurants.

3.1.1 Dataset Structure Overview

The dataset consists of three key components: Sentences, Aspects, and Polarity. The Sentences provide insights into actual user reviews for restaurants, containing information on various aspects. The Aspects categorize specific parts of the service or product being reviewed, such as food, service, staff, price, ambience, menu, place, and miscellaneous, allowing for targeted sentiment analysis. The Polarity component quantifies the sentiment as positive, negative, or neutral, helping to understand the sentiment behind the reviews. The model aims to classify the sentiment of the input sentences into one of these three polarities, capturing the overall vibe conveyed in the reviews.

3.1.2 Dataset Analysis

The Aspect Bar Chart shows the distribution of different aspects in restaurant reviews. The y-axis lists aspects like Food, Staff, Miscellaneous, Place, Service, Menu, Ambience, and Price, while the x-axis shows the number of mentions. Food is the most mentioned aspect with over 2,000 mentions, followed by Staff with around 1,500. Other aspects have fewer mentions, highlighting Food and Staff as predominant aspects. The Polarity Bar Chart displays sentiment polarities in reviews. The y-axis categorizes sentiments into Positive, Neutral, and Negative, and the x-axis shows the number of mentions. Neutral sentiments are the most common with around 3,000 mentions, followed by Negative with 2,000, and Positive as the least common. This indicates that most reviews are neutral or negative. Below is an overview of the aspects and polarities shown in the bar charts. in **Figure 5** and **Figure 6**.

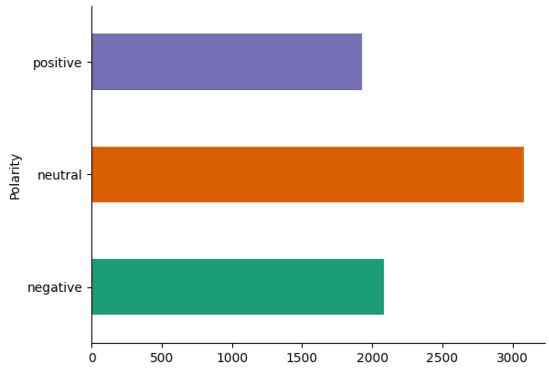

Figure 5: Bar Chart of Polarity

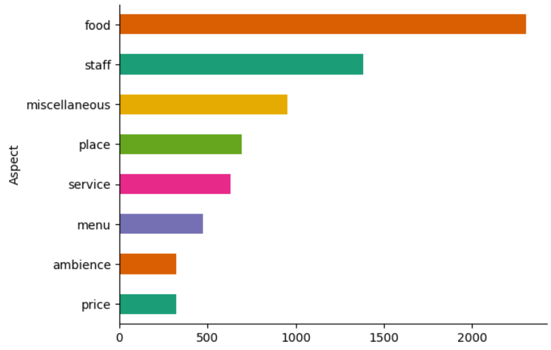

Figure 6: Bar Chart of Aspect

3.2 Experiment Setup

Table 1 below shows the set of values we used in hyperparameter tuning is listed followed by final hyperparameters that we used in the final models,

Parameter Names	Values
number of layers	(1, 2, 3)
learning rate	(0.1, 0.01, 0.001)
dropouts	(0.1, 0.2, 0.3, 0.4, 0.5)
hidden size	(64, 128, 256, 512)

(a) Range of Hyperparameter Values

Parameter Names	Values
number of layers	2
learning rate	0.01
dropouts	0.5
hidden size	128

(b) Final Hyperparameter Values

Table 1: Hyperparameter Tuning: Range of Values Tested and Final Values Chosen

In considering the multiple loss and optimization methods we came to conclusion that Adam Optimizer and Cross Entropy Loss is working better for our models than other methods.

3.2.1 Text Preprocessing methods

The preprocessing stage of our models take the textual data and performs several key steps to prepare it for analysis or model training. First, it removes punctuation from each sentence and replaces it with spaces. Next, it tokenizes the sentences into individual words and converts all words to lowercase to ensure uniformity. Lemmatization is applied to reduce words to their base forms. After preprocessing, the text data is transformed into embeddings using pre-trained GloVe vectors ("glove-wiki-gigaword-50"), which provide numerical representations of words based on their semantic meaning therefore, cleaned, normalized, and converted into a format suitable for machine learning models.

3.2.2 Experimentation with stop words

We tried fitting the models by using different strategies such as keeping stopwords in the sentences, excluding stopwords from the sentences and to customize the stopwords to be removed, in which we kept negative words like no, never, not etc. in the sentence and therefore fit the model. we found out that the model is better when we used customized stopwords. The reason for that is we observed that negative words have a lot of information with them when it comes to aspect based sentimental analysis.

4 Results

We present quantitative and qualitative results to evaluate the performance of our models.

4.1 Quantitative Results

The Table 2 below shows the model variants and its corresponding performance accuracies

Models	Training Accuracy	Test Accuracy	Precision	Recall	F1 Score
Variant 1	86.83%	63.37%	61.91%	63.37%	61.75%
Variant 2	75.02%	63.60%	62.74%	63.60%	62.58%
Variant 3	81.68%	64.04%	64.33%	64.04%	64.14%

Table 2: Model variants and their performance accuracies

Results in above table totally justifies the structure we used. we can see variant 3 is performing well than others because we are using attention mechanism into that as referred by [**4**] and also by the works done in [**3**], which is giving weightage to the relevant words in a sentence according to aspect. The variant 3 not only has high accuracy but also has good weighted precision, Recall and F1 score.

4.1.1 Ablation Study

The Table 3 below shows other Seq2Seq model experimentations and its corresponding performance accuracies.

Models	Training Accuracy	Test Accuracy
Bi-RNN with Attentions	72.91%	59.49%
Bi-GRU with Attentions	86.90%	62.93%
Bi-RNN with Attentions but appending aspect with each word	76.30%	57.38%
Bi-GRU with Attentions but appending aspect with each word	80.94%	64.59%

Table 3: Other Seq2Seq models with performance accuracies

The above ablation study in table 3 clearly shows the attention models are performing well than without attention neural networks(table 2). This was expected as well because attention mechanism gives high importance to the relevant words according to aspect. We also observed that when we append aspect with each word embeddings then it does not give much improvement and GRU are outperforming RNN which was expected as well.

4.2 Qualitative Results

As an instance from the test data we are considering the sentence **"The atmosphere was wonderful, however the service and food were not"** with aspects as **food** and **service** and polarity of both the aspects are **negative**. The figure below illustrates the words in the test sentence with different attention weights.

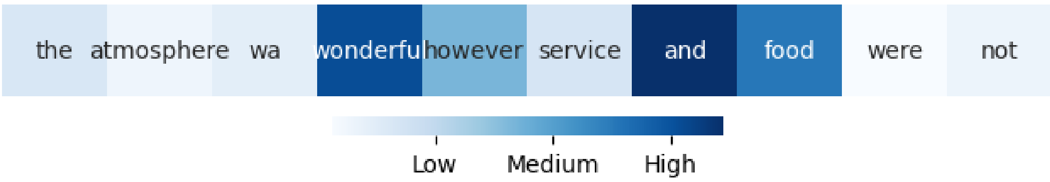

Figure 7: Attention Weights Heat Map for aspect as **food**

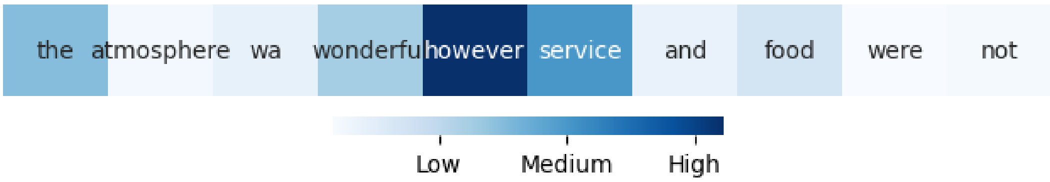

Figure 8: Attention Weights Heat Map for aspect as **service**

We can see from the above heat map that relevant words related to aspect are getting more attention weight in the attention model, such as for aspect as Service, the words like service, however , not, wonderful are getting attention. Whereas for aspect as Food, the words like "food", "and", "not","wonderful" are having more weights.

5 Conclusion

In our study, we investigated the effectiveness of Bi-RNN models with various approaches to incorporating aspect information, including attention mechanisms. Our findings indicate that attention-based models consistently outperform other architectures, such as Bi-RNN and Bi-GRU with attention, as observed in our ablation study. We were successful in achieving good sentimental classification model specifically attention based. However, we noted the quality of dataset controls upto a great level your classification models performances and also noted limitations in the dataset, particularly in the imbalance of occurrences for the "Neutral" polarity and the "Food" aspect. Furthermore, we observed that model performance varied depending on the structure or method of aspect integration. To address these issues, future research could focus on refining aspect utilization strategies and exploring the potential benefits of incorporating transformers with attention mechanisms to further enhance model performance.

References

[1] C. Zhang, L. Liu, B. Liu, and T. Liu, "Research on semantic sentiment analysis based on BiLSTM," in *2021 4th International Conference on Artificial Intelligence and Big Data*, 2021.

[2] Graves, Jaitly, Navdeep and Mohamed, Abdel-rahman, "Hybrid speech recognition with deep bidirectional LSTM," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, 2013, pp. 273–278, doi: 10.1109/ASRU.2013.6707742.

[3] V. R. Kota and S. D. Munisamy, "High accuracy offering attention mechanisms based deep learning approach using CNN/bi-LSTM for sentiment analysis," *International Journal of Intelligent Computing and Cybernetics*, vol. 15, pp. 61–74, 2022.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[5] Y.-H. Hsieh and X.-P. Zeng, "Sentiment analysis: An ERNIE-BiLSTM approach to bullet screen comments," *Sensors*, vol. 22, no. 14, p. 5223, 2022, doi: 10.3390/s22145223.